DOCUMENT RESUME

ED 177 173 TM 009 511

AUTHOR Haertel, Edward H.; And Others
TITLE Metric-Free Distributional Comparisons.
INSTITUTION Central Midwestern Regional Educational Lab., Inc., Chicago, Ill. ML-GROUP for Policy Studies in Education.
PUB DATE May 78
NOTE 33p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Comparative Statistics; *Comparative Testing; Goodness of Fit; *Statistical Analysis; Test Interpretation
IDENTIFIERS *Negative Hypergeometric Distribution

ABSTRACT

Two methods are presented for comparing distributions, such as achievement test score distributions, for distinctly different groups of persons in such a way that the comparison will not be influenced by the particular metric of the test being used. Both methods use percentile scores. One method, attributed to Flanagan, fits a straight line to the graph representing the two distributions on normalized score scales. The second method finds the negative hypergeometric score distributions for each of the two groups that most closely approximate their observed score distributions. It has been shown that many observed test score distributions closely approximate the negative hypergeometric distribution with the same mean, variance, and number of items. These smoothed distributions are used as the basis of comparison between groups. An example is given in which the Flanagan line closely approximates the relationship shown by smoothed curves relating the distributions of two groups, where the median for one group was above the 93rd percentile of the other group. (CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED177173

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Metric-Free Distributional Comparisons*

Edward H. Haertel

William A. Thrash

David E. Wiley

May, 1978

ML-GROUP for Policy Studies in Education

CEMREL, Inc.

875 North Michigan Avenue

Suite 3311

Chicago, Illinois 60611

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Edward H. Haertel
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM009 511

* This paper is Appendix A to Social and Economic Differences in Elementary School Achievement.

The concept of educational disadvantage -- whether defined in terms of available resources, educative activities, or knowledge and skill -- has been a relative one. That is, there is a perceived "gap" between the disadvantaged and those who are better off. In articulating and deepening this conception through empirical investigation, methods are needed which describe these gaps in realistic and concise fashion.

The social concepts which enter into this discussion are primarily one which apply to groups of individuals -- blacks, Hispanics, dropouts, the poor, etc. Thus, it is the differences between such groups and more advantaged ones which are the gaps studied in documenting and explaining their disadvantages. When group characteristics are thought to be measured quantitatively, i.e., characterized with numbers having interpretations of "more" and "less", they are usually summarized as distributions. These distributions attach absolute or relative frequencies to each specific "quantity" of the characteristic. Differences in the relative frequencies -- proportions of the groups -- corresponding to particular values reflect group differences or "gaps" in the measured characteristic.

These group differences are usually analyzed in two stages:

1) The separate distributions are summarized using statistics such as means, standard deviations, medians, or other quantiles;

2) These statistics are compared. The disadvantages of this procedure are that:

   a) The summary statistics may not be equally informative for all of the distributions to be compared;

   b) There is usually no way to describe and compensate for such differential adequacy; and

   c) If the metric of the characteristic to be compared is arbitrary or specific to a particular measuring instrument, the resulting group contrasts are difficult to compare with those resulting from other instruments addressing other characteristics.

In this study we decided to begin with another approach. Instead of summarizing the distributions and comparing the summaries, we have attempted to compare the distributions and summarize the comparisons. We have found this strategy to have several benefits.

1) It is possible to compare entire distributions in a manner that is free of the metric of the original measuring instrument.

2) It is feasible to summarize -- usually with a small set of values -- these comparisons in a fashion which allows:

   a) a description of the adequacy of these summaries; and

b) their augmentation if the original set is inadequate.

3) In applications to educational and psychological test data these procedures yield distributional summaries -- means and standard deviations -- that are traditional in form yet free of the original measurement metrics.

1. THE RATIONALE

The primary methodological objective of this study was to describe differences in entire distributions of achievement for various groups. This was to be accomplished in an easily interpretable way that would permit comparisons of the magnitude of these differences for different tests and grades.

Sample means and standard deviations are dependent on the original metrics of the various tests. Traditional methods of test equating do deal with the comparison of entire distributions, but focus on the transformation of test data from one metric to another and require equating tables relating the metrics of more than one test. Latent trait models may one day be useful in drawing such comparisons, but have not yet been developed in sufficient generality for our purposes.

A feasible solution. Drawing on the strategy outlined above, we may sketch a solution to this problem. The cumulative distributional information for two populations with respect to a single test can be summarized by the number triples $(x, p_1(x), p_2(x))$, where x ranges over the score points on the test, $p_1(x)$ is the percentile value for score point x in group 1, and $p_2(x)$ is the percentile value for that score point in group 2. This information can be displayed conveniently as the graph of $p_1(x)$ vs. $p_2(x)$, which can be formally defined as that collection of ordered pairs $(p_1(x), p_2(x))$ for x in the range of the test.

These pairs of percentiles are invariant under any monotone transformation of the test score X. As almost all reasonable equating methods are monotonic transformations of the observed scores (1), we have in the ordered pairs (p_1, p_2) an instrument-free summary of the groups' performances on the test.

To give a common basis for comparison, we can choose one group and use it as a reference group to which others are compared. The reference group might be, for instance (as in the example of Section V.), the collection of all non-poor white pupils whose mothers graduated from high school. We might also choose to make our entire sample the reference group, as has been done in the main body of this report.

While the paired percentiles are completely informative and metric-free, they are not easily interpretable in terms of group differences. We can

facilitate interpretation by finding a standard and interpretable metric based solely on the paired percentile values. In fact, given that we have defined reference and comparison groups, we may accomplish this by fixing a standard distribution for the reference group (2). In what follows we have arbitrarily chosen the unit normal. The primary rationale for this choice is that the mean and variance of a normal distribution are among the most fundamental and easily understood of statistical concepts.

This fixing of a reference distribution specifies a distribution function relating the scale values in the reference metric to the reference percentiles. As the reference percentiles are paired with those of the comparison group, this also fixes a comparison distribution in the same scale as the reference distribution. The form of the comparison distribution is not at all constrained to be similar to that of reference distribution chosen as the standard. In practice, however, our experience has been that transforming the metric to make the reference group value normal generally serves to (approximately) normalize those in the comparison group as well. The mean and standard deviation of these transformed scores can then be compared to the corresponding values of 0 and 1 for the reference group. Moreover, this summary of the differences between distributions is free of the original metric of the test, therefore can be compared across grades and tests. This is the basis for the interpretations in this report.

The following heuristic argument describes the procedure in greater detail. (It is not a proof, as matters of domain and existence of inverses are ignored.)

Let F and G, respectively, be the distribution functions (c d f) of scores in the reference and comparison groups. Our basic information is the set of ordered pairs (F (t), G (t)) for raw scores t. Assume that there is a metric in which Group 1 is normal, i.e., that there is a transformation

$$u = h(t) \quad \text{such that}$$

$$F(t) = Fh^{-1}(h(t)) = Fh^{-1}(u) = \Phi(u)$$

where Φ is the unit normal cdf. Rescaling the metric in the same way for the comparison group gives

$$G(t) = Gh^{-1}\, h(t) = Gh^{-1}(u)$$

and, since

$$Gh^{-1} = Fh^{-1}\, hF^{-1}\, Gh^{-1} = \Phi\, hF^{-1}\, Gh^{-1},$$

we may let

$$k = hF^{-1}\, Gh^{-1}$$

to give

$$Gh^{-1}(u) = Fh^{-1}(k(u)) = \Phi(k(u)).$$

Since $\Phi(u)$ is normal $(0, 1)$, we know that $\Phi(k(u))$ is normal (μ, σ^2) if and only if $k(u) = \sigma u + \mu$, that is, if the graph $(u, k(u))$ of u vs. $k(u)$ is a straight line with slope σ and intercept μ.

This rather complicated discussion become useful when we note that

$$\Phi^{-1}(f(t)) = \Phi^{-1}(\Phi(u)) = u$$

and

$$\Phi^{-1}(G(t)) = \Phi^{-1}(\Phi(k(u))) = k(u).$$

This implies that by plotting

$$\Phi^{-1}(F(t)) \quad \text{vs} \quad \Phi^{-1}(G(t))$$

we can judge the extent to which the comparison group has been normalized by evaluating how closely this graph approximates a straight line, and can estimate the rescaled mean and standard deviation for the comparison group from the slope and intercept of a fitted straight line.

It would be possible to find a suitable straight line by ordinary or weighted least squares regression of

$$\Phi^{-1}(u) \quad \text{on} \quad \Phi^{-1}(k(u)),$$

but there is a simpler procedure, a modification of one proposed by Flanagan (4), which eliminates much of the calculation. The graph of

$$\Phi^{-1}(u) \quad \text{vs.} \quad \Phi^{-1}(k(u))$$

will always cross the vertical axis at the point

$$(0, \Phi^{-1}(m_2))$$

and the horizontal at

$$(\Phi^{-1}(m_1), 0)$$

where m_1 is the percentile in group 1 corresponding to the median in group 2 and m_2 is the percentile in group 2 corresponding to the median in group 1 (that is, the points $(m_1, .5)$ and $(.5, m_2)$ on the original graph). Thus we may use $\Phi^{-1}(m_2)$ as our estimate of μ and $-\Phi^{-1}(m_2) / \Phi^{-1}(m_1)$ as our estimate of σ. In the discrete case when there is no point exactly on the axis, interpolation can be used. Other quantiles could replace the medians in this procedure, with appropriate modifications, but the means would not work, as the Φ^{-1} transformation preserves the order information of the quantiles, but not the metric information of the means.

2. SMOOTHING

Although the procedure described above yields summary statistics which are interpretable and free from the original test metric, sampling fluctuations can lead to highly variable results for small cell sizes. A regression procedure might be slightly more stable than the Flanagan procedure, but would not likely be satisfactory for many of the small cell sizes we face. We have chosen to solve this problem by using prior information about the likely shapes of test score distributions to smooth observed data.

Lord & Novick (5, pp. 515 - 521) present strong theoretical arguments and empirical evidence that observed test scores can, in many situations similar to ours, be closely approximated by negative hypergeometric distributions,

$$h(x) = P[X = x] = \frac{(b+n-1)(b+n-2)\cdots b(-n)(-n+1)\cdots(-n+x-1)\; a(a+1)\cdots(a+x-1)}{(a+b+n-1)(a+b+n-2)\cdots(a+b)(-b-n+1)(-b-n+2)\cdots(-b-n+x)\; x!}$$

where n is the number of items on the test and a and b are parameters to be determined. By using the method of moments, the parameters a and b can easily be estimated from the mean μ, variance σ^2, and Kuder-Richardson formula 21 reliability α_{21} of the observed score distribution as

$$a = \mu(1-\alpha_{21})/\alpha_{21} \qquad \text{and}$$

$$b = -a + n(1-\alpha_{21})/\alpha_{21}$$

where n is the number of items on the test. Moreover, α_{21} itself can be determined from only μ, σ^2 and n as

$$\alpha_{21} = \frac{n}{n-1}\left(1 - \frac{\mu(n-\mu)}{n\sigma^2}\right) \quad (5, \text{p. } 517)$$

This result allows us to smooth distributions of observed scores with ease, even for small cell sizes. The procedure described in Section I for normalizing pairs of distributions and interpreting differences between them can now be applied to pairs of fitted observed distributions. The scaled means and standard deviations from pairs of fitted observed distributions are the basic descriptive statistics of our report, and form the basis of much of the discussion in the text.

Corresponding to the fitted negative hypergeometric model for observed scores, Lord and Novick propose a beta distribution of true scores,

$$g(\xi) = \xi^{a-1}(1-\xi)^{b-1}/B(a,b),$$

where a and b are the same parameters as in the negative hypergeometric and

$$B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$$

is the usual beta function. These fitted true score distributions lead to the same conclusions as the fitted observed distributions, and are therefore largely ignored in the body of the report.

A compound binomial model (5) might fit the observed data better than the binomial-negative hypergeometric model used in this report, but the amount of computation required would be prohibitive, and the increased number of parameters in this more general model would likely lead to overfitting of our small cells. We only need the model to reduce the effect of sampling fluctuations in the data - with large amounts of data no explicit model of test scores would be needed. The compound binomial model has enough parameters to provide an excellent fit to exactly those sampling fluctuations we are trying to smooth away. For these reasons the binomial model is a better choice.

3. ACCUMULATION AND AGGREGATION

After examing in great detail the procedures described above it was decided not to focus, for interpretive purposes, primarily on the direct summaries based on the observed percentile values, but to first smooth the observed distributions and use the resulting percentile values for subsequent computation. This greatly eased the burden of accumulation and aggregation by making it possible to carry only the mean, standard deviation, and number of items for each group, rather than the complete score distribution. It also greatly reduced sampling errors, for most cells.

Individuals' scores (appropriately weighted) were accumulated within cells formed by combining levels of selected characteristics, e.g., hours spent doing homework, mother's education, number of children in family. The fitting and scaling procedures described above were then carried out for each cell in a table.

4. THE ALGORITHM

Given the number of items, n, and the means and variances μ_R and σ_R^2 of the reference distribution and μ_c and σ_c^2 of the comparision distribution,

1) Find the KR21 reliability

$$\alpha_{21} = \frac{n}{n-1}\left(1 - \frac{\mu(n-\mu)}{n\sigma^2}\right)$$

for the reference distribution.

2) Use this to find the parameters

$$a = \mu(1-\alpha_{21})/\alpha_{21}$$

and

$$b = -a + n(1-\alpha_{21})/\alpha_{21}$$

of the negative hypergeometric distribution.

3) Repeat (1) and (2) for the comparison distribution.

4) Generate the negative hypergeometric distribution $h(x)$, $x = 0,1, \ldots, n$ for the reference distribution.

5) Similarly compute $k(x)$, the negative hypergeometric approximation to the comparison distribution.

6) Sum to find the cdf's $H(x) = \sum_{i=0}^{x} h(i)$

and $K(x) = \sum_{i=0}^{x} k(i)$

7) Find the median of the fitted comparison distribution. In this case we found x_L such that

$$K(x_L) \leq .5 \leq K(x_L + 1)$$

and solved for the median x_* by linear interpolation, as $x_* = x_L + (.5 - K(x_L)) / (K(x_L + 1) - K(x_L))$

Observe that if $K(x_L) = .5$ then $x_* = x_L$ is an integer.

8) Find $m_2 = H(x_*)$ by interpolation as

$$m_2 = H(x_L) + (x_* - x_L)\ (H(x_L + 1) - H(x_L))$$

Observe that if $x_* = x_L$ is an integer then $m_2 = H(x_L)$.

9) $\Phi^{-1}(m_2)$ is the scaled mean of the comparison group. Φ^{-1} can be found in a probit table or from subroutines for the inverse unit normal available on most computers (and many calculators).

10) Find $m_1 = K(y_*)$ where y_* is the median of the reference distribution. The procedure is exactly as in (7) - (9), but with H and K interchanged throughout.

11) $-\Phi^{-1}(m_2) / \Phi^{-1}(m_1)$ is the scaled standard deviation of the comparison group.

5. EXAMPLE

Let us take as our example reading comprehension below-grade-level for grade 1, taking non-poor whites whose mothers were high school graduates as the reference group and poor blacks whose mothers did not finish high school as the comparison group. The observed and smoothed probability distributions for these groups are displayed in Figure 1. The raw cumulative percentiles are given in Table 1.

FIGURE 1. Reading Comprehension Score Distributions (Density Functions) (Below-level Test) for Reference (Non-poor, white, maternal high school graduate) and Comparison (poor, black, high school dropout) Groups, First Grade.

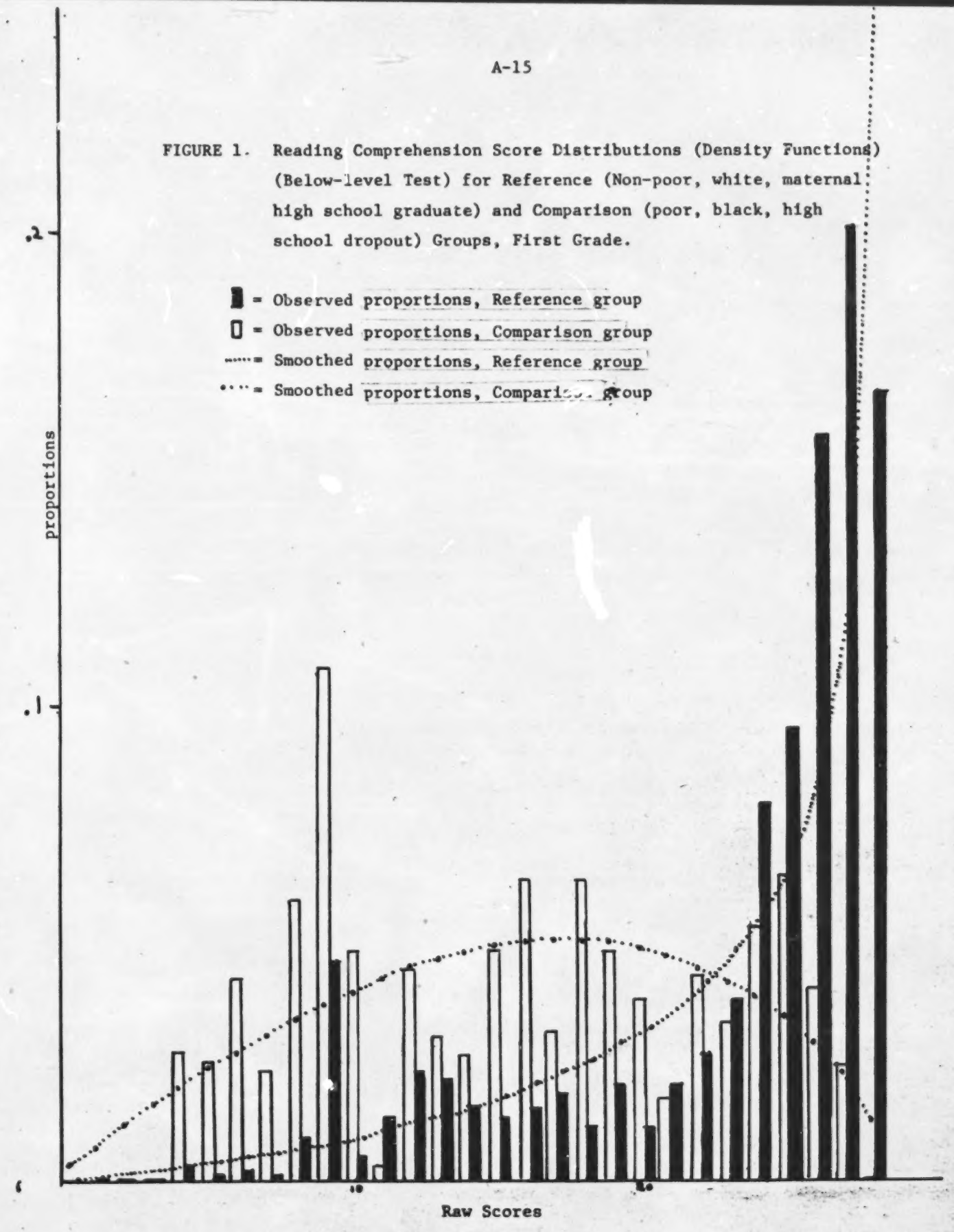


TABLE 1. Observed and Smoothed Cumulative Distributions, and Their Differences (Residuals) for the Reference and Comparison Groups.

	OBSERVED		SMOOTHED		RESIDUALS	
SCORE	REFERENCE	COMPARISON	REFERENCE	COMPARISON	REFERENCE	COMPARISON
0	0.0	0.0	0.000	0.004	-0.000	-0.004
1	0.001	0.0	0.001	0.011	-0.001	-0.011
2	0.001	0.0	0.003	0.023	-0.002	-0.023
3	0.001	0.0	0.005	0.039	-0.004	-0.039
4	0.004	0.027	0.008	0.058	-0.004	-0.031
5	0.005	0.052	0.012	0.082	-0.007	-0.030
6	0.007	0.094	0.017	0.109	-0.011	-0.015
7	0.008	0.117	0.023	0.140	-0.015	-0.023
8	0.017	0.176	0.030	0.174	-0.013	0.002
9	0.053	0.284	0.038	0.211	0.015	0.073
10	0.056	0.332	0.047	0.251	0.010	0.082
11	0.071	0.335	0.058	0.293	0.013	0.042
12	0.094	0.379	0.071	0.338	0.024	0.041
13	0.115	0.409	0.085	0.384	0.030	0.025
14	0.130	0.435	0.100	0.432	0.029	0.003
15	0.143	0.483	0.118	0.481	0.025	0.002
16	0.158	0.546	0.138	0.531	0.020	0.015
17	0.176	0.577	0.161	0.581	0.016	-0.004
18	0.187	0.640	0.186	0.631	0.001	0.009
19	0.207	0.688	0.215	0.681	-0.007	0.008
20	0.216	0.726	0.247	0.729	-0.031	-0.003
21	0.238	0.743	0.283	0.776	-0.045	-0.033
22	0.264	0.786	0.325	0.821	-0.061	-0.035
23	0.302	0.819	0.374	0.863	-0.071	-0.044
24	0.381	0.872	0.431	0.902	-0.050	-0.030
25	0.476	0.936	0.500	0.936	-0.024	0.000
26	0.633	0.976	0.589	0.965	0.045	0.012
27	0.834	1.000	0.715	0.987	0.119	0.013
28	1.000	1.000	1.000	1.000	0.0	0.0

Figure 2 is a graph of these cumulative percentiles, with percentiles in the comparison group as abscissa and those in the comparison group as ordinate. The diagonal line in the display is the line of equality. I.e., the points would fall on this line if the distributions were identical. The fact that all of the plotted points lie on a (moderately) smooth curve below and to the right of this line implies that the comparison distribution is, at every reference percentile value, shifted toward higher percentile values. That is, poor black pupils whose mothers are high school dropouts have uniformly lower achievement than non-poor, white pupils whose mothers graduated.

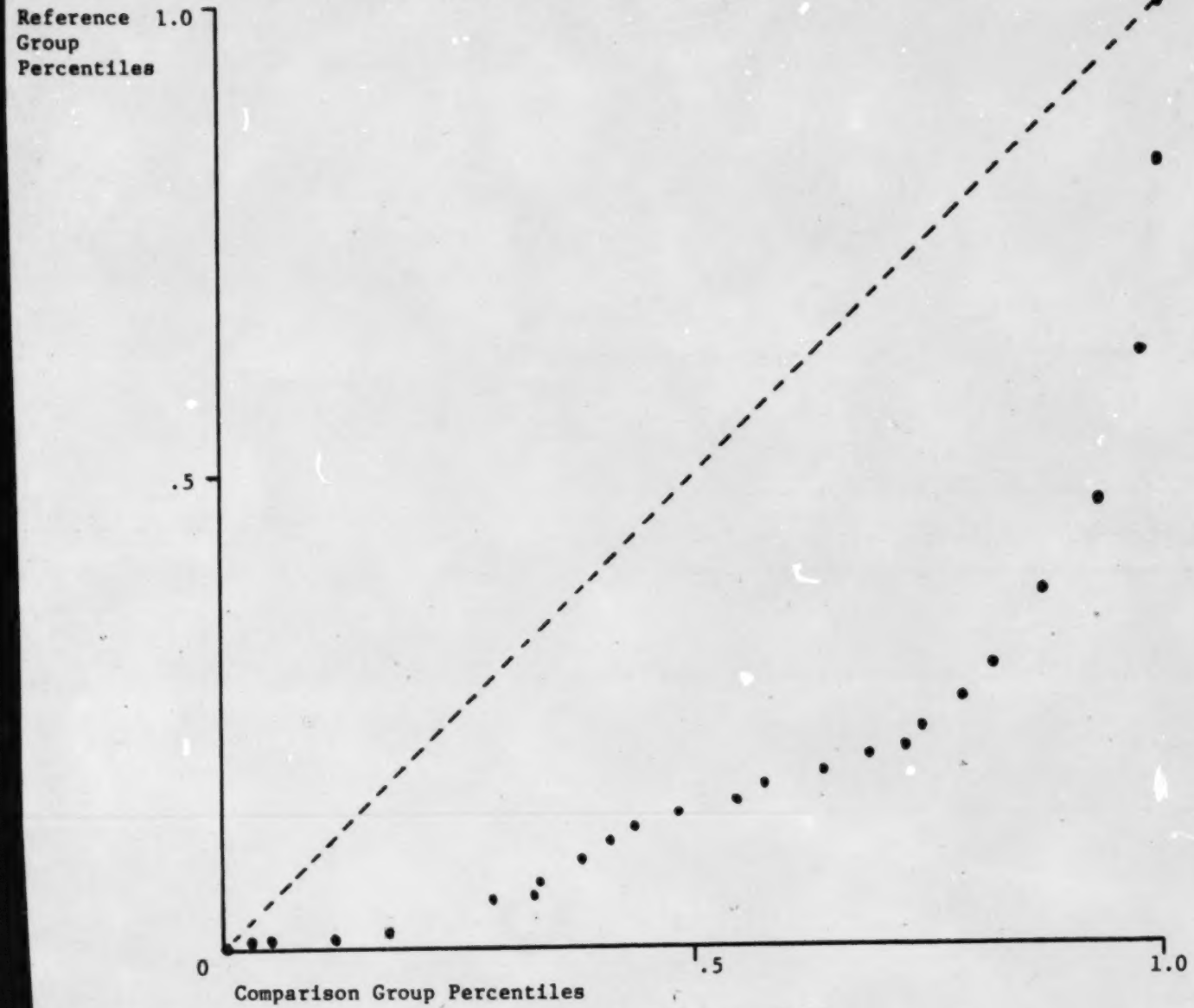
Table 2 contains $\Phi^{-1}(p_1)$ and $\Phi^{-1}(p_2)$ in columns 1 and 2. These normal deviates are plotted in Figure 3. If we can simultaneously normalize scores by both groups, this graph should be roughly a straight line. Though there is some evidence of systematic deviation from normality, the plot is remarkably linear, especially since these groups were chosen to have an extremely large discrepancy in the parameters of their fitted distributions.

TABLE 2. Inverse Normal Transformed Values (Probits) of the Observed and Smoothed Cumulative Percentiles, and Their Differences (Residuals) for the Reference and Comparison Groups.*

	OBSERVED		SMOOTHED		RESIDUALS	
SCORE	REFERENCE	COMPARISON	REFERENCE	COMPARISON	REFERENCE	COMPARISON
0	******	******	-3.343	-2.685	******	******
1	-3.239	******	-2.996	-2.282	-0.243	******
2	-3.239	******	-2.754	-1.997	-0.485	******
3	-3.014	******	-2.560	-1.766	-0.454	******
4	-2.656	-1.921	-2.396	-1.568	-0.260	-0.353
5	-2.546	-1.627	-2.250	-1.392	-0.295	-0.235
6	-2.484	-1.316	-2.119	-1.231	-0.365	-0.085
7	-2.417	-1.190	-1.997	-1.080	-0.421	-0.110
8	-2.116	-0.931	-1.883	-0.938	-0.233	0.007
9	-1.618	-0.570	-1.774	-0.803	0.156	0.233
10	-1.576	-0.433	-1.670	-0.672	0.094	0.239
11	-1.469	-0.427	-1.570	-0.544	0.101	0.118
12	-1.315	-0.309	-1.471	-0.419	0.157	0.110
13	-1.200	-0.231	-1.375	-0.295	0.175	0.064
14	-1.127	-0.165	-1.279	-0.172	0.152	0.007
15	-1.067	-0.043	-1.184	-0.048	0.117	0.005
16	-1.004	0.115	-1.088	0.077	0.085	0.038
17	-0.929	0.194	-0.991	0.204	0.063	-0.011
18	-0.889	0.357	-0.893	0.334	0.003	0.023
19	-0.816	0.491	-0.791	0.469	-0.025	0.021
20	-0.785	0.600	-0.685	0.610	-0.100	-0.010
21	-0.711	0.652	-0.573	0.759	-0.138	-0.107
22	-0.631	0.792	-0.453	0.919	-0.178	-0.127
23	-0.517	0.912	-0.322	1.094	-0.195	-0.182
24	-0.302	1.134	-0.174	1.291	-0.128	-0.157
25	-0.059	1.522	0.000	1.520	-0.059	0.001
26	0.341	1.982	0.224	1.806	0.117	0.176
27	0.968	******	0.568	2.216	0.400	******
28	******	******	******	******	******	*******

*Asterisked values correspond to transformations of percentile values of 0.0 and 1.0 or their differences.

FIGURE 2. The (Paired) Percentile Plot, Reference vs. Comparison Group



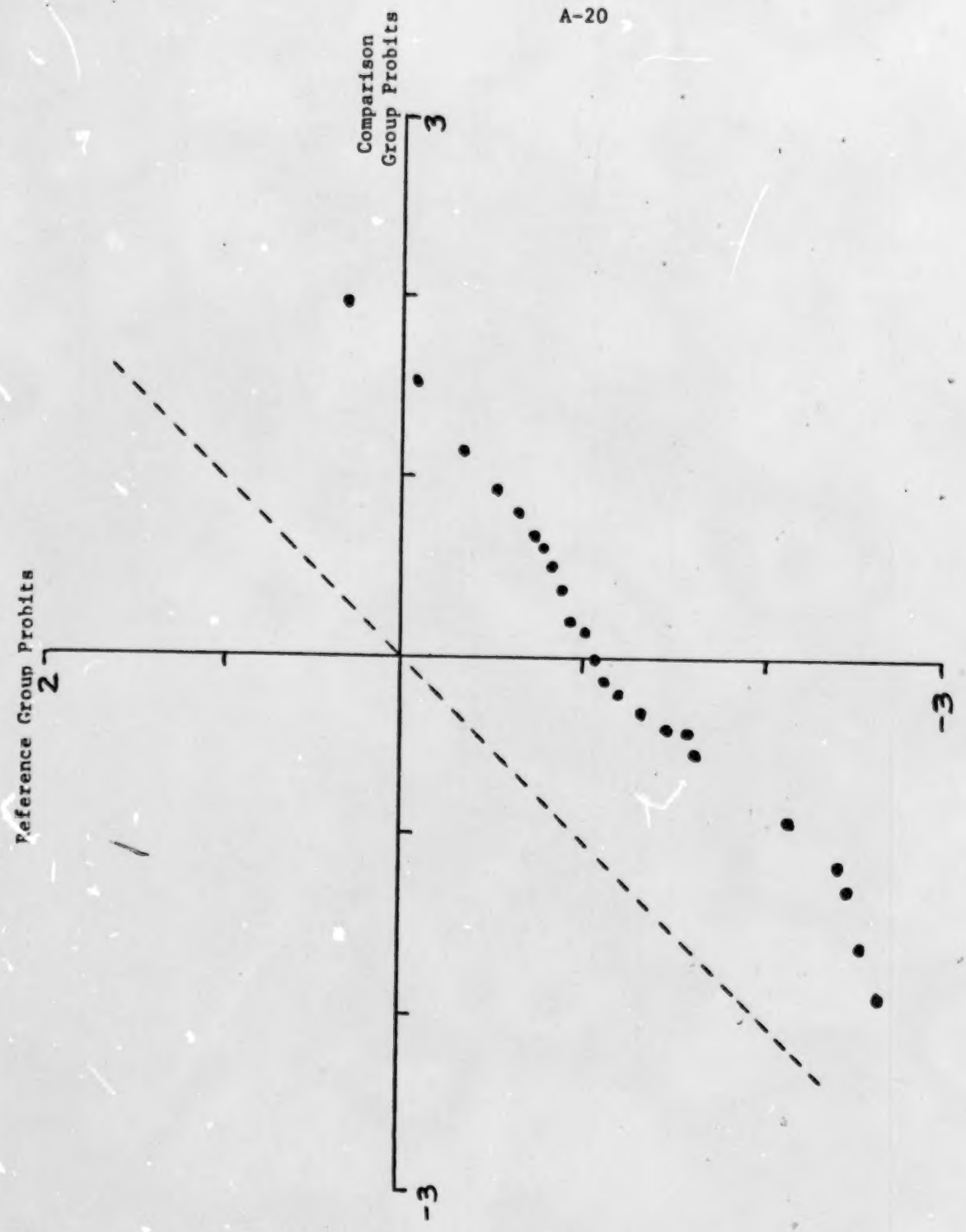


Figure 3. Inverse-Normal Transformed Values (Probits) of Paired Percentiles, Reference vs. Comparison Group.

Figure 4 shows graphically how the Flanagan line could be fitted to these raw data. The two short solid lines represent graphically the linear interpolation to find m_1 and m_2, the dashed line is the Flanagan line. We can see that this is not really a very good fit to the data, but even so we can compute the scaled mean = -1.05 and scaled standard deviation = 0.64.

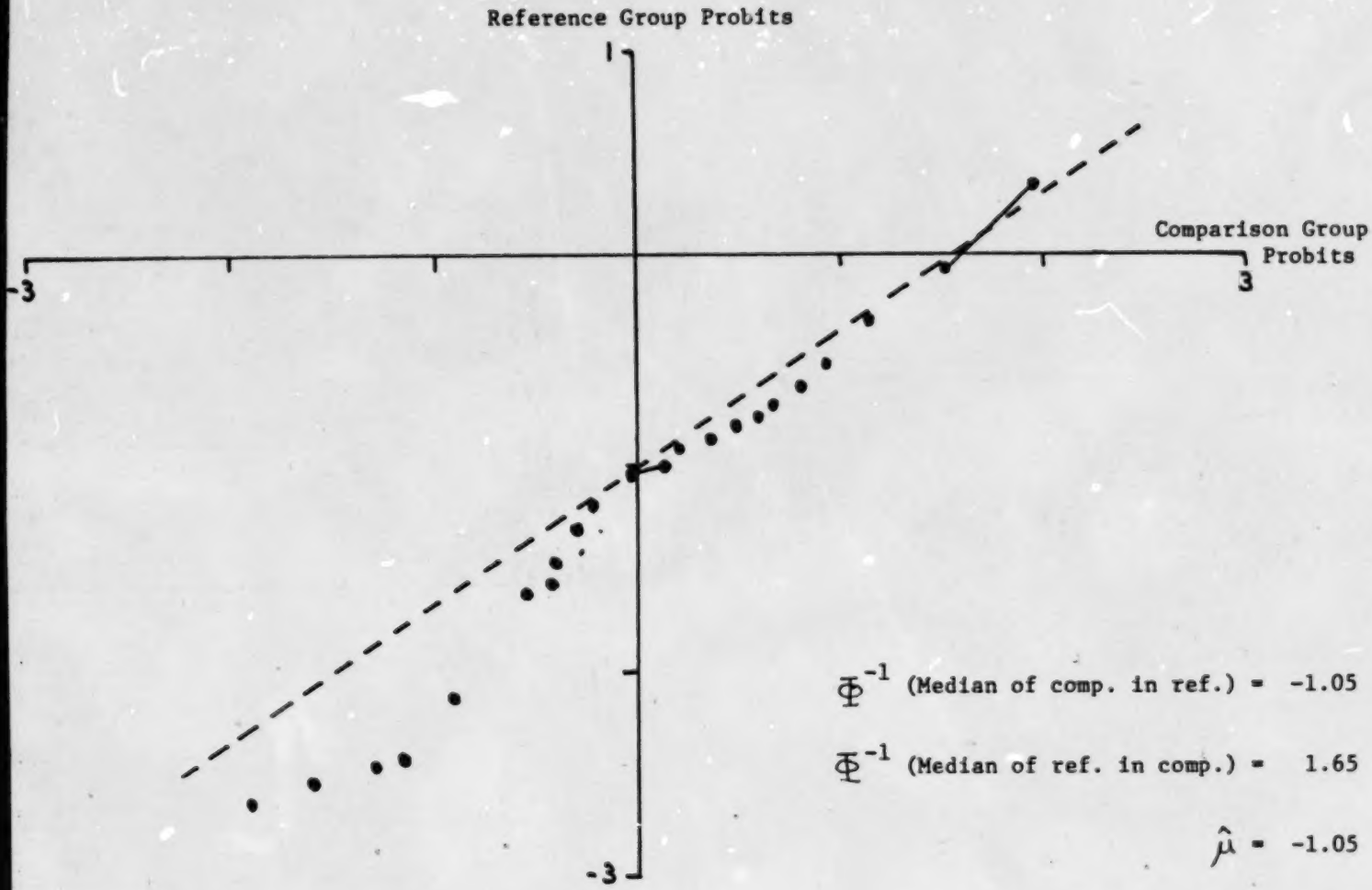


FIGURE 4. Linear Fit to Probit Plot of the Observed Percentiles Using the Flanagan Procedure

Let us now compute the parameters of the fitted distributions:

For the reference group $\mu = 23.219$, $\sigma^2 = 32.781$ and $n = 28$ so

$$\alpha_{21} = (28 - 23.219\ (28 - 23.219) / 32.781) / 27 = .9116$$

from which

$a = 23.219\ (1 - .9116) / .9116 = 2.2513$ and

$b = -2.2513 + 28\ (1 - .9116) / .9116 = .4635.$

Repeating the process for the comparison group we find

$a = 2.1540$ and

$b = 1.7197.$

The cumulative percentiles of these hypergeometric distributions were given in Table 1.

The corresponding probits are shown on columns 3 and 4 of Table 2 and are plotted along with the raw values in Figure 4. Certainly the fitted distribution is better behaved than the raw! The Flanagan line, based on the smoothed values, is also shown in Figure 5. This line is an excellent fit to the smooth data, and is actually a better fit to the original data points than the Flanagan line computed from those points.

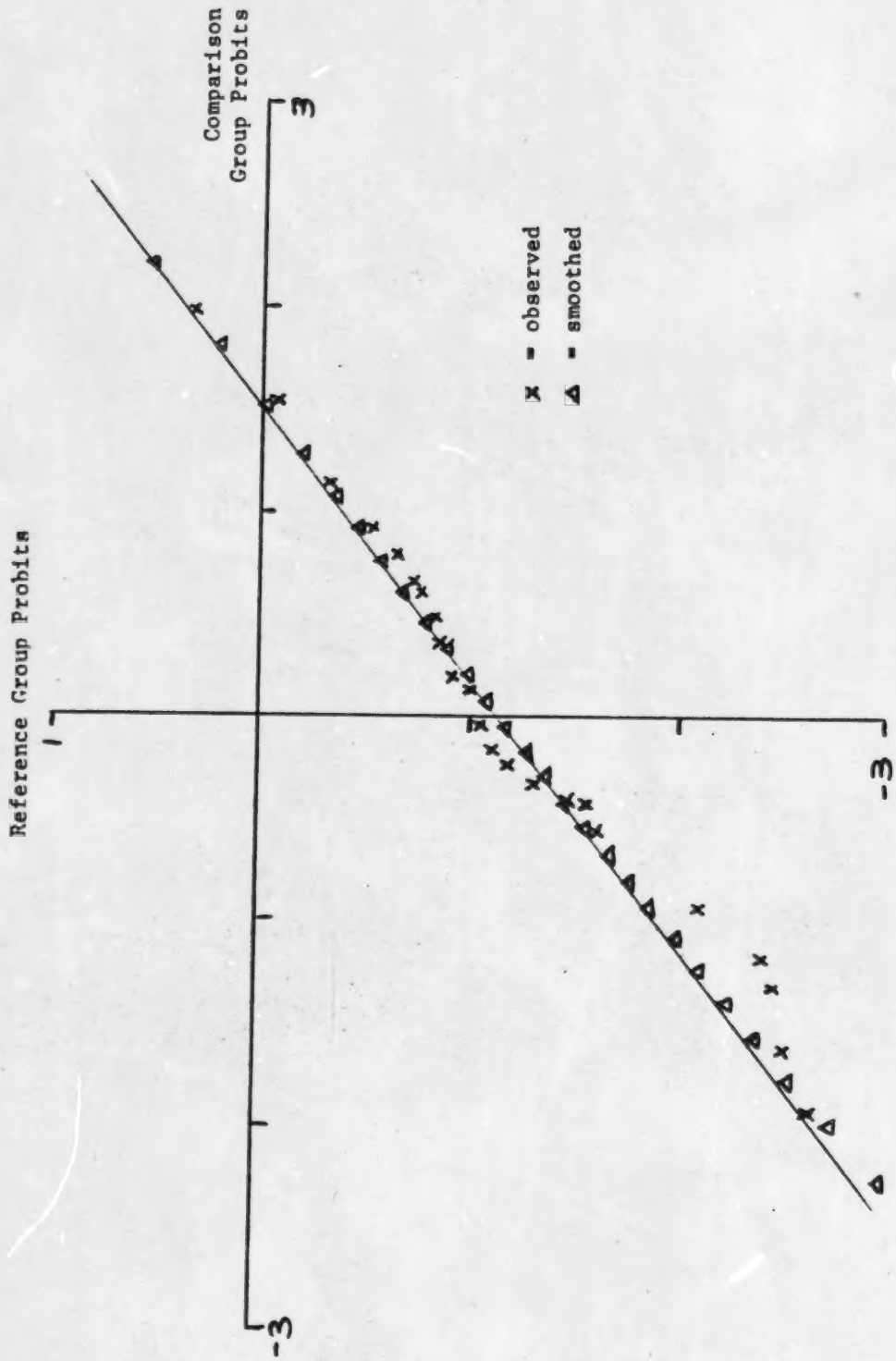


Figure 5. Linear Fit to Probit Plot of Smoothed Percentiles Using the Flanagan Procedure.

The computations for this fitted line are quite simple: In the comparison group K(15) = .481 and K(16) = .531

and 15 + (.5 - .481)/(.531 - .481) = 15.38,

but

H(15) = .118 and H(16) = .138,

so

m_2 = H (15.38) = .118 + (15.38 - 15)(.138 - .118) = .126

and the scaled mean is

$\Phi^{-1}(.126) = -1.146.$

The other median is even easier to find:

m_1 = K(25) = .936, and $\Phi^{-1}(.936) = 1.522,$

so the scaled deviation is

-(-1.146/1.520) = .754.

One more version of the data can be seen in Table 3 and Figure 6. Figure 6 is really Figure 5 (minus one point in the extreme left tail) from a different angle; the graph has been rotated clockwise 45 degrees and reflected in the horizontal axis. The intent of this rotation is to make the (equality) baseline horizontal. This baseline becomes the horizontal axis of the graph and the distance between the fitted line and the axis at the origin is the scaled comparison mean. If the scaled (comparison)standard devistion is one, the fitted line is also horizontal. The result is easier for the eye to grasp; the interpretation is the same.

TABLE 3. Inverse Normal Transformed Values (Probits) of Observed and Smoothed Percentiles, for the Reference and Comparison Groups.*

	Raw		Fitted Observed		Residuals	
Score	$\frac{C+R}{2}$	$\frac{C-R}{2}$	$\frac{C+R}{2}$	$\frac{C-R}{2}$	$\frac{C+R}{2}$	$\frac{C-R}{2}$
0	******	******	-3.014	.329	******	******
1	******	******	-2.639	.357	******	******
2	******	******	-2.376	.379	******	******
3	******	******	-2.163	.397	******	******
4	-2.289	.368	-1.982	.414	-.307	-.047
5	-2.087	.460	-1.821	.429	-.265	.030
6	-1.900	.584	-1.675	.444	-.225	.140
7	-1.804	.614	-1.539	.459	-.266	.156
8	-1.524	.593	-1.411	.473	-.113	.120
9	-1.094	.524	-1.289	.486	.195	.039
10	-1.005	.572	-1.171	.499	.167	.073
11	- .948	.521	-1.057	.513	.110	.009
12	- .812	.503	- .945	.526	.134	-.024
13	- .716	.485	- .835	.540	.120	-.056
14	- .646	.481	- .726	.554	.080	-.073
15	- .556	.511	- .616	.568	.061	-.056
16	- .445	.560	- .506	.583	.062	-.024
17	- .368	.562	- .394	.598	.026	-.037
18	- .266	.623	- .280	.614	.013	.010
19	- .163	.654	- .161	.630	-.002	.023
20	- .093	.693	- .038	.648	-.055	.045
21	- .030	.682	.093	.666	-.123	.016
22	.081	.712	.233	.686	-.153	.026
23	.198	.715	.386	.708	-.189	.007
24	.416	.718	.559	.733	-.143	-.015
25	.732	.791	.760	.760	-.029	.030
26	1.162	.821	1.015	.791	.147	.030
27	******	******	1.392	.824	******	******
28	******	******	******	******	******	******

* Asterisked values correspond to transformations of percentile values of 0.0 and 1.0 or their differences.

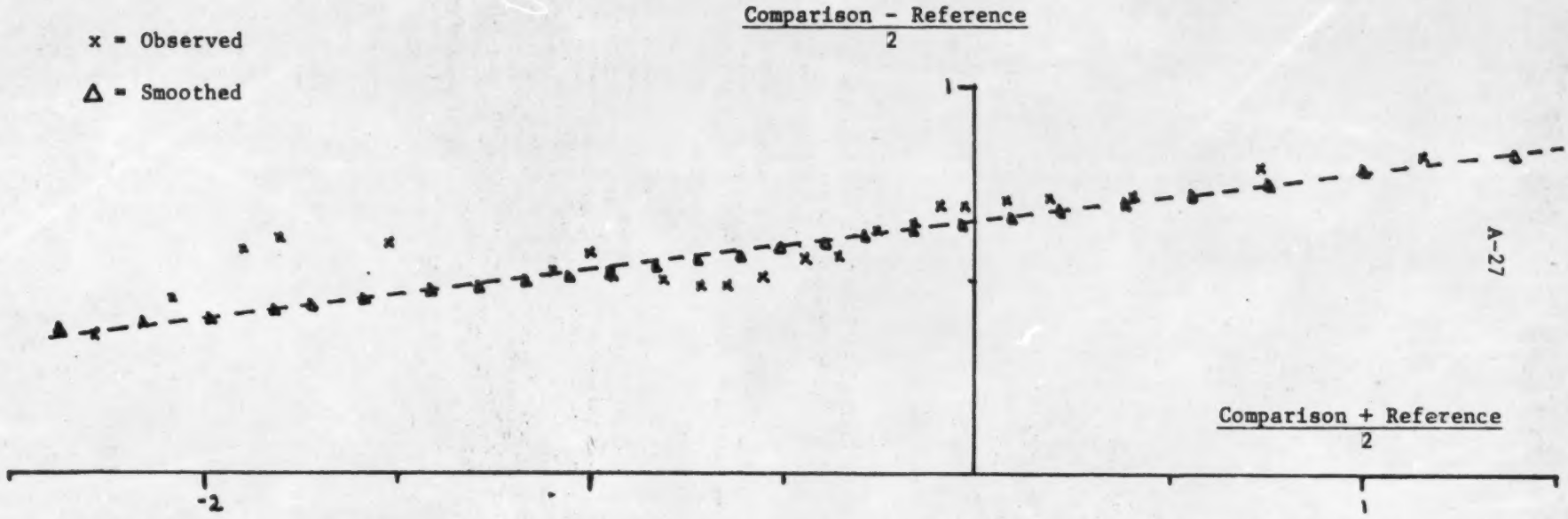


FIGURE 6. Linear Fit to Probit Plot of Smoothed Percentiles Using the Flanagan Procedure, Rotated Plot

NOTES

1. At least when items are scored correctly.

2. We can approximate the discrete distributions of scores by continuous distributions, if we choose.

3. Fh^{-1} denotes function composition, i.e., $Fh^{-1}(u) = F(h^{-1}(u))$

4. See, for instance, W. H. Angoff, "Scales, Norms, and Equivalent Scores", in R. L. Thorndike (ed.), Educational Measurement, Second Edition. (Washington, D.C.: American Council on Education, 1971).

5. F. M. Lord and M. R. Novick, Statistical Theories of Mental Test Scores (Reading, Mass.: Addison-Wesley Publishing Company, 1968). Our notation differs slightly from theirs, particularly in that Lord and Novick's b is our b plus n-1.